SILVERTREE
HOLDINGS

How Data Science can be applied in the Online Shopping Industry

Case Study: PriceCheck Product Feed Process Optimization

Francois van Heerden

Silvertree Holdings
www.silvertree.holdings

eCommerce in South Africa

Revenue (US$)
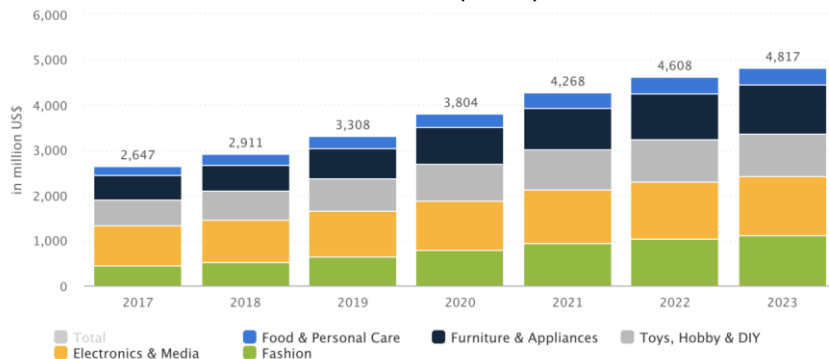
Revenue Growth Rate (%)
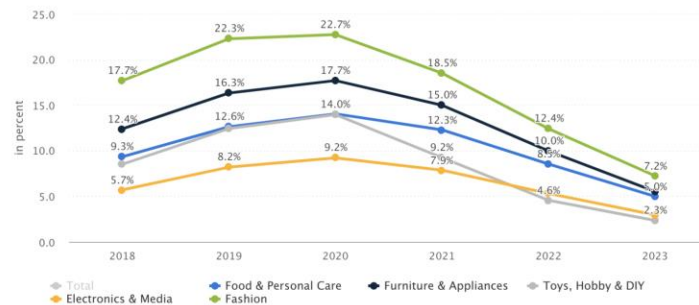
Age Distribution

Payment & Device

Source: Statista Feb 2019

SILVERTREE HOLDINGS
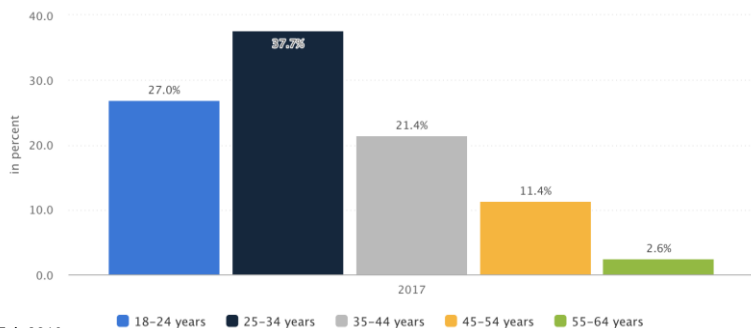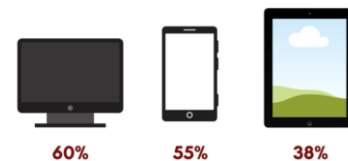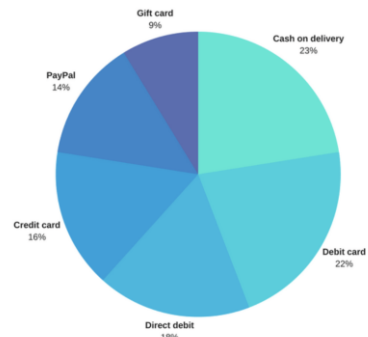
Online vs Physical Retail

ONLINE

- Digital
- Track every interaction by user/ customer
- Personalised shopping
- Customers can switch to a new online store with very little effort
- Can't experience the product before delivery
- Delivered to your door or click and collect

VS

PHYSICAL

- Brick and Mortar Location
- High level data
- Average customer model
- Loyalty can be retained by locational convenience
- Physical interaction with product before purchase
- Collect items from store

SILVERTREE
HOLDINGS

Online Data Science Focus Areas

DATA/ AI/ ML

Process Efficiency

Customer Experience

Productivity/ Content/ Marketing

Self-Service Analytics

Personalised Shopping Experience

Customer Portfolio Management
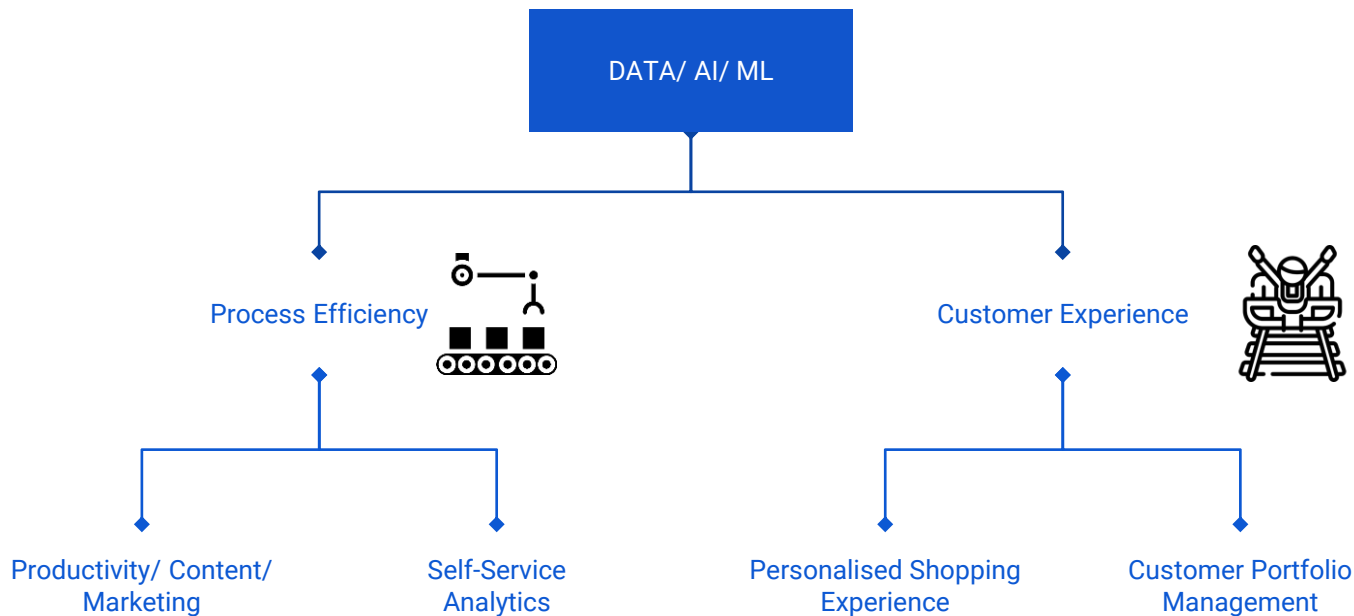
Models/Methods

- Recommendation Engines
- Customer Segmentation
- Dynamic Pricing
- Behavioural Predictions
- Chat Bots
- Etc.

SILVERTREE
HOLDINGS

Case Study: PriceCheck

Problem: Categorical miss classification within search and category navigation on the website.

Example:

Categories > Toys 🤔

KKONION 2 Pcs Covers L Shaped Sofa Universal Elastic Stretch Corner Sofa Covers Modern Sectional...

☆ ☆ ☆ ☆ ☆ 0 Reviews   Write a Review

The Default Delivery Time is About 10-15 Business Days,Usually Get the Items in About 12 days.2 Pcs Covers for L Shaped Sofa Universal

Brand: Kkonion   Category: Educational

Add to compare

Why there isn't a quick fix by humans?

⏰

6 million products -> 8500 Hours

☀️

+-1060 working days

Backlog of 49 months x 50 months to complete

= 2450 months or 200 years 👴👵

Case Study: Current System

Product Feed

Map Categories to PriceCheck

Verify Classification

Stores import products into Shopping Feeder. The system receives information regarding current name, category tree, description etc.

Challenges:
*Missing Data
*Low Quality Data

Use the product feed metadata to map the categories to PriceCheck categories.

Challenges:
*Misclassification
*Don't know which classification has a high probability of misclassification

Content team needs to work through the products to verify the classification. Then only will the category be changed.

Challenges:
*Millions of products per month
*Bias to work on most popular products

SILVERTREE
HOLDINGS

Proposed Solution: NLP Machine Learning Classification

DATA → PREPROCESSING → MODEL → PRODUCTION
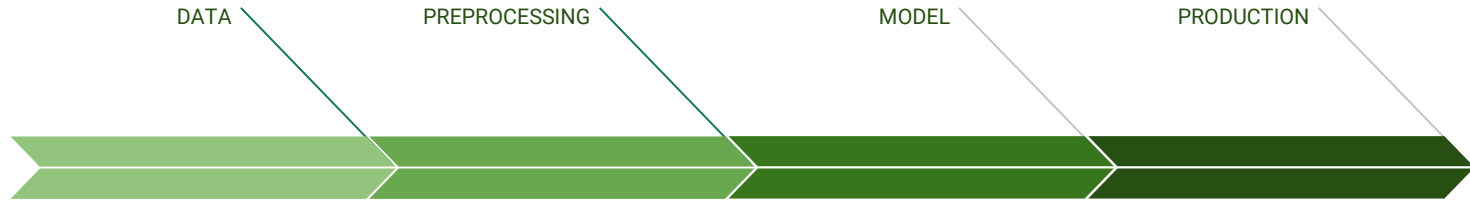
PRODUCT METADATA

Create structured product level data including categorical data.

TEXT CLEANING

Python function to remove punctuation and stop words. Create binary labels.
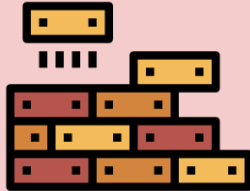
NLP Multinomial NB

Designed for classification with discrete features like word counts for text classification.

API

Integrate the model into product classification system.

Challenges:

- Data quality
- Scale
- Business trust of AI

```
70 #train the model
71 Y = df['b_labels'].values
72
73 count_vectorizer = CountVectorizer(decode_error='ignore')
74 X = count_vectorizer.fit_transform(df['data'])
75
76 # split up the data
77 Xtrain, Xtest, Ytrain, Ytest = train_test_split(X, Y, test_size=0.33)
78
79 #run model learning
80 model = MultinomialNB()
81
82 model.fit(Xtrain, Ytrain)
83 print("train score:", model.score(Xtrain, Ytrain))
84 print("test score:", model.score(Xtest, Ytest))
```

SILVERTREE
HOLDINGS

Business Impact

01 — 1000's of labour hours saved
- Avoid increasing labour cost
- Increase labour productivity
- Team will focus on more strategic work

02 — Improved Customer Experience
- Improved search functionality
- Improve overall product offering (navigation and search)
- Improve offer comparison

03 — Increase in overall business efficiency
- More productive workforce focussing on key performance areas that grow the business in the long run
- Case Study for feature AI/ML projects

SILVERTREE HOLDINGS

THANK YOU!